PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2002-268681
(43)Date of publication of application : 20.09.2002

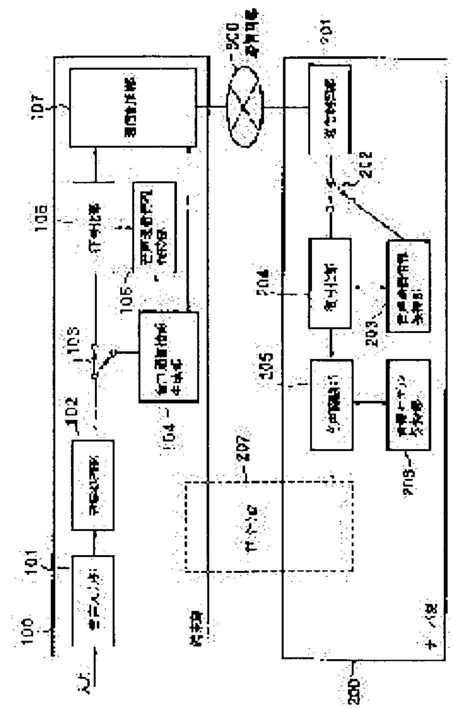
(51)Int.Cl. G10L 15/28
G10L 15/06
G10L 15/00
G10L 19/00

(21)Application number : 2001-065383
(22)Date of filing : 08.03.2001

(71)Applicant : CANON INC
(72)Inventor : KOSAKA TETSUO
YAMAMOTO HIROKI

(54) SYSTEM AND METHOD FOR VOICE RECOGNITION, INFORMATION PROCESSOR USED FOR THE SAME SYSTEM, AND METHOD THEREOF

(57)Abstract:
PROBLEM TO BE SOLVED: To enable proper encoding corresponding to variations of acoustic features and to prevent a decrease in recognition rate due to variations of environmental sound and a decrease in compressibility resulting from the encoding.
SOLUTION: In a terminal part 100, a sound processing part 102 analyzes sound information inputted through a sound input part 101 to obtain multi- dimensional feature quantity parameters. At initial setting time, a sound communication information generation part 104 sets processing conditions for compressive encoding according to the multi-dimensional feature quantity parameters and the processing conditions are held in a sound communication information holding part 105 and further held in a sound communication information holding part 203 of a server part 200. In voice recognition, an encoding part 106 compresses and encodes the feature quantity parameters obtained by the sound processing part 102 under the processing conditions held in the sound communication information holding part 105 and a decoding part 204 of the server part 200 decodes the parameters under the processing conditions held in the sound communication information holding part 203.



LEGAL STATUS
[Date of request for examination]
[Date of sending the examiner's decision of rejection]
[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]
[Date of final disposal for application]
[Patent number]
[Date of registration]
[Number of appeal against examiner's decision of rejection]
[Date of requesting appeal against examiner's decision of rejection]
[Date of extinction of right]

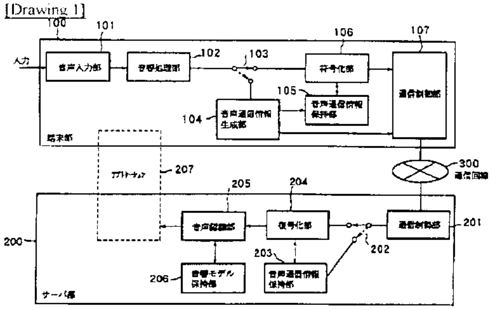
* NOTICES *

JPO and NCIPI are not responsible for any damages caused by the use of this translation.

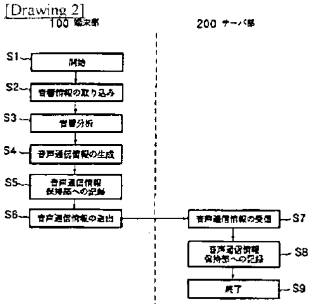
1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

DRAWINGS

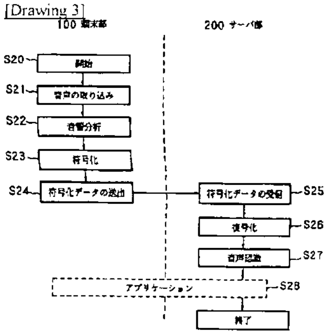
[Drawing 1]



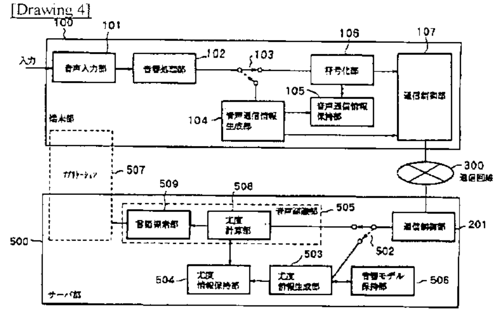
[Drawing 2]



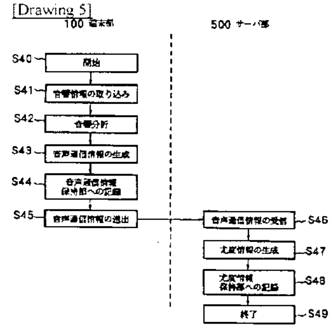
[Drawing 3]



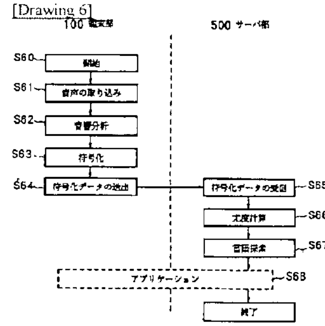
[Drawing 4]



[Drawing 5]



[Drawing 6]



[Translation done.]

DESCRIPTION OF DRAWINGS

[Brief Description of the Drawings]
[Drawing 1] It is the block diagram showing the speech recognition structure of a system in the 1st operation gestalt.
[Drawing 2] It is a flow chart explaining initial setting of the voice recognition system in the 1st operation gestalt.
[Drawing 3] It is a flow chart explaining speech recognition processing of the voice recognition system in the 1st operation gestalt.
[Drawing 4] It is the block diagram showing the speech recognition structure of a system in the 2nd operation gestalt.
[Drawing 5] It is a flow chart explaining initial setting of the voice recognition system in the 2nd operation gestalt.
[Drawing 6] It is a flow chart explaining speech recognition processing of the voice recognition system in the 2nd operation gestalt.

[Translation done.]

DETAILED DESCRIPTION

[Detailed Description of the Invention]
[0001]
[Field of the Invention] This invention relates to a voice recognition system and equipments, and those approaches.
[0002]
[Description of the Prior Art] In recent years, the attempt used as an input interface of a device is made with the advance of a speech recognition technique. When using a speech technology as an input interface, it is common to introduce the configuration for speech processing into the device concerned, to perform speech recognition within the device concerned, and to treat this as alter operation to the device concerned.
[0003] On the other hand, development of a small personal digital assistant in recent years enabled the small personal digital assistant to realize many processings. Sufficient input key cannot be provided from the constraint on the size of a small personal digital assistant. Fo this reason, requests of wanting to use the above speech recognition techniques for the operator guidance for realizing various functions are mounting.
[0004] As one method of realizing this, there is a method of carrying a speech recognition engine in the small personal digital assistant itself. However, in such a small personal digital assistant, resources, such as memory and CPU, are restricted in many cases, and a highly efficient recognition engine may be unable to be carried. Then, a small personal digital assistant is connected with a server in networks, such as wireless, and the speech recognition of the client-server mold which performs the part with little processing cost on a terminal, and performs a part with much throughput by the server is proposed among speech recognition processings.
[0005] In this case, since little direction is desirable, as for the amount of data transmitted to a server from a terminal, transmitting, after compressing data (coding) is common. Moreover, not the general voice coding approach that is used with the cellular phone but the coding approach suitable for sending the data about speech recognition is proposed about the coding approach for it.
[0006]
[Problem(s) to be Solved by the Invention] In coding suitable for speech recognition used in the speech recognition of the client-server mold mentioned above, after asking for an audio feature parameter generally, the approach of encoding these parameters by the formation of a scalar quantity child, or vector quantization or subband quantization is taken. And especially about the acoustical description at the time of carrying out speech recognition, do not take coding in this case into consideration, and it is made.
[0007] However, speech recognition is used under a noise environment, or the optimal coding processings also differ in the case where the property of the microphone used in the case of speech recognition differs from a general thing. For example, since distribution of the feature parameter of the voice under a noise environment the case of the above-mentioned approach differs from distribution of the feature parameter of the voice under a quiet environment, it is desirable for the range of quantization to be also adapted for it.
[0008] For the ********** reason, in the former, the technical problem that the compressibility in the case of degradation of a recognition rate and coding could not be enlarged occurred coding under the noise environment, without taking change of the above acoustical descriptions into consideration.
[0009] This invention is made in view of the above-mentioned technical problem, suitable coding according to change of the acoustical description is enabled, and it aims at preventing decline in the recognition rate by change of an environmental sound, and decline in the compressibility in coding.
[0010]
[Means for Solving the Problem] The voice recognition system by this invention for attaining the above-mentioned purpose is equipped with the following configurations. Namely, an input means to input sound information and an analysis means to analyze the sound information inputted with said input means, and to acquire a characteristic quantity parameter, A setting means to set up the processing conditions for compression coding based on the characteristic quantity parameter obtained with said analysis means, It has the recognition means which carries out speech recognition based on the conversion means which carries out compression coding of the characteristic quantity parameter obtained with said analysis means according to said processing conditions, the processing conditions set up with said setting means, and the characteristic quantity parameter by which compression coding was carried out with said conversion means.
[0011] Moreover, the speech recognition approach by this invention for attaining the above-mentioned purpose is equipped with the following configurations. Namely, the input process which inputs sound information and the analysis process which analyzes the sound information inputted at said input process, and acquires a characteristic quantity parameter, The setting process which sets up the processing conditions for compression coding based on the characteristic quantity parameter obtained at said analysis process, It has the recognition process which carries out speech recognition based on the conversion process which carries out compression coding of the characteristic quantity parameter obtained at said analysis process according to said processing conditions, the processing conditions set up at said setting process, and the characteristic quantity parameter by which compression coding was carried out at said conversion process.
[0012] Moreover, the information processor of this invention for attaining the above-mentioned purpose An input means to input sound

information, and an analysis means to analyze the sound information inputted with said input means, and to acquire a characteristic quantity parameter, A setting means to set up the processing conditions for compression coding based on the characteristic quantity parameter obtained with said analysis means, It has a 1st notice means to notify the processing conditions set up with said setting means to an external device, the conversion means which carries out compression coding of the characteristic quantity parameter obtained with said analysis means according to said processing conditions, and a 2nd notice means to notify the data obtained with said conversion means to said external device.
[0013] Moreover, the information processor for attaining the above-mentioned purpose A 1st receiving means to receive the processing conditions which start compression coding from an external device, A maintenance means to make the processing conditions received with said 1st receiving means hold in memory, It has a 2nd receiving means to receive the data by which compression coding was carried out from said external device, and a recognition means to perform speech recognition to the data received with said 2nd receiving means using the processing conditions held at said maintenance means.
[0014] Moreover, the information processing approach for attaining the above-mentioned purpose The input process which inputs sound information, and the analysis process which analyzes the sound information inputted at said input process, and acquires a characteristic quantity parameter, The setting process which sets up the processing conditions for compression coding based on the characteristic quantity parameter obtained at said analysis process, It has the 1st notice process which notifies the processing conditions set up at said setting process to an external device, the conversion process which carries out compression coding of the characteristic quantity parameter obtained at said analysis process according to said processing conditions, and the 2nd notice process which notifies the data obtained at said conversion process to said external device.
[0015] Furthermore, the information processing approach for attaining the above-mentioned purpose The 1st receiving process which receives the processing conditions which start compression coding from an external device, The maintenance process which makes the processing conditions received at said 1st receiving process hold in memory, It has the 2nd receiving process which receives the data by which compression coding was carried out from said external device, and the recognition process which performs speech recognition to the data received at said 2nd receiving process using the processing conditions held at said maintenance process.
[0016]
[Embodiment of the Invention] Hereafter, the gestalt of suitable operation of this invention is explained with reference to an accompanying drawing.
[0017] <1st operation gestalt> drawing 1 is the block diagram having shown the speech recognition structure of a system concerning the 1st operation gestalt. Moreover, drawing 2 and drawing 3 are the flow charts explaining actuation of the voice recognition system indicated by the block diagram of drawing 1 . Hereafter, including an example of operation, drawing 1 , drawing 2 , and drawing 3 are related, and it explains.
[0018] In drawing 1 , 100 is the terminal section and can apply the various personal digital assistants containing a cellular phone etc. 101 is the voice input section, with a microphone etc., incorporates a sound signal and digitizes this. 102 is the acoustical-treatment section, performs sonagraphy and generates a multi-dimension sound parameter. In addition, a mel cepstrum, a delta mel cepstrum, etc. can use the analysis method generally used for speech recognition for sonagraphy. 103 is the processing change-over section and switches data flow by the processing at the time of initialization later mentioned with reference to drawing 2 and drawing 3 , and the processing at the time of speech recognition.
[0019] 104 is the voice communication information generation section, and generates the data for encoding to the sound parameter obtained in the acoustical-treatment section 102. With this operation gestalt, the voice communication information generation section 104 clusters the data of each dimension of a sound parameter in the class (referred to as 16step(s) in this example) of arbitration, and generates a clustering resulting table using the result divided by clustering. 105 is a voice communication information attaching part, and holds the clustering resulting table generated in the voice communication information generation section 104. In addition, as a record medium which makes a clustering resulting table hold in the voice communication information attaching part 105, various things, such as memory, such as RAM, a floppy (trademark) disk (FD), and a hard disk (HD), can be used.
[0020] 106 is the coding section and encodes the multi-dimension sound parameter obtained from the acoustical-treatment section 102 using the clustering resulting table recorded on the voice communication attaching part 105. 107 is the communications control section and sends out a clustering resulting table, the encoded sound parameter to a communication line 300.
[0021] 200 is the server section and performs speech recognition about the encoded multi-dimension sound parameter which is sent from the terminal section 100. The usual personal computer etc. can constitute the server section 200.
[0022] 201 is the communications control section and receives the data transmitted from the communications control section 107 of the terminal section 100 through the circuit 300. 202 is the processing change-over section and switches data flow by the processing at the time of initialization later mentioned with reference to drawing 2 and drawing 3 , and the processing at the time of speech recognition.
[0023] 203 is a voice communication information attaching part, and holds the clustering resulting table which received from the terminal section 100. In addition, as a record medium for making a clustering resulting table hold in the voice communication information attaching part 203, various things, such as memory, such as RAM, a floppy disk (FD), and a hard disk (HD), can be used.
[0024] 204 is the decryption section and decodes the coded data (multi-dimension sound parameter) received from the terminal section 100 in the communications control section 201 with reference to the clustering resulting table held at the voice communication information attaching part 203. 205 is the speech recognition section and performs recognition processing about the multi-dimension sound parameter obtained in the decryption section 204 using the sound model held at the sound model attaching part 206.
[0025] 207 is application and performs various processings based on the result of speech recognition. Application 207 may be performed by the server section 200 side, and may be performed by the terminal section 100 side. However, in the case of the application performed by the terminal section 100 side, it is necessary to notify an audio recognition result to the terminal section 100 through the communications control sections 201 and 107.
[0026] In addition, at the time of initialization, the processing change-over section 103 of the terminal section 100 switches connection to the voice communication information generation section 104 so that data may flow to the coding section 106 at the time of speech recognition processing. Similarly, at the time of initialization, the processing change-over section 202 of the server section 200 switches

connection to the voice communication information attaching part 203 so that data may flow to the decryption section 204 at the time of speech recognition processing. These processing change-over sections 103 and 202 interlock and operate. The change has two kinds of modes, for example, initial learning mode and recognition mode, and when initial learning mode is directed in order to learn before recognition use with directions of a user, as for the processing change-over section 103, data switch connection so that the processing change-over section 202 may flow to the voice communication information attaching part 203 to the voice communication information generation section 104. Since a user directs recognition mode when actually recognizing, it is answered, and as for the processing change-over section 103, as for the processing change-over section 202, data switch connection to the decryption section 204 to the coding section 106 so that it may flow.
[0027] In addition, 300 is a communication line which connects the terminal section 100 and the server section 200, and its various means of communications which can transmit data are available irrespective of a cable and wireless. In addition, each part of the terminal section 100 mentioned above and the server section 200 is realized by performing the control program with which CPU which each has was stored in memory. Of course, hardware may realize the part or all the functions of each part.
[0028] Hereafter, the flow chart of drawing 2 and drawing 3 explains the actuation in the above-mentioned voice recognition system to a detail.
[0029] Initial setting shown in the flow chart of drawing 2 before speech recognition initiation is performed first. In initial setting, the coding conditions for fitting coded data to a sound environment are set up. Although coding and speech recognition of voice data are possible by using the default created based on the sound condition in a quiet environment even if this initial setting does not carry out for example, improvement in a recognition rate is expected by performing initial setting.
[0030] In initial setting, first, in step S2, sound data are incorporated in the voice input section 101, and A/D conversion is performed further. The sound data inputted are the voice data at the time of uttering voice by acoustical environment similar to the acoustical environment or it which is actually used. The effect of the property of the used microphone is also reflected in this sound data. Moreover, when there is a noise generated inside a background or a device, it becomes data influenced [the].
[0031] At step S3, sonagraphy of the sound data inputted in the voice input section 101 is performed in the acoustical-treatment section 102. As mentioned above, sonagraphy can use the analysis method generally used for speech recognition for a mel cepstrum, a delta mel cepstrum, etc. As mentioned above, since the processing change-over section 103 connects the voice communication information generation section 104 side at the time of initialization, creation of the data for coding processing is performed by the voice communication information generation section 104 in step S4.
[0032] The creation approach of the data performed in this voice communication information generation section 104 is explained below. About coding for speech recognition, the formation of a scalar quantity child, vector quantization, the approach of carrying out subband quantization, etc. are used in it in quest of a sound parameter. Especially in this operation gestalt, although it is not necessary to limit the technique and any approach can be used, here explains the approach by the formation of a scalar quantity child. By this approach, each dimension of the multi-dimension sound parameter called for by the sonagraphy of step S3 is formed into a scalar quantity child. To the formation of a scalar quantity child, various approaches are possible.
[0033] Two examples are shown below.
1) LBG generally used as the approach:clustering technique by LBG -- use law. the data of each dimension of a sound parameter -- LBG -- it divides into the class (for example, 16step(s)) of arbitration using law.
2) How to assume a model : the data of each dimension of a sound parameter assume that Gaussian distribution is followed. Within the limits of 3sigma of the whole distribution of each dimension is divided so that it may become an area division-into-equal-parts rate, i.e., same probability, for example, it clusters to 16step(s).
[0034] Furthermore in step S6, the clustering resulting table called for in the voice communication information generation section 104 is transmitted to the server section 200. If in charge of a transfer, a clustering resulting table is transmitted to the server section 200 using the communications control section 107 in the terminal section 100, a communication line 300, and the communications control section 201 of the server section 200.
[0035] In the server section 200, the communications control section 201 receives a clustering resulting table in step S7. At this time, the processing change-over section 202 will have connected the voice communication information attaching part 203 and the communications control section 201, and the clustering resulting table received by the voice communication information attaching part 203 at step S8 will be recorded.
[0036] Next, the processing at the time of speech recognition is explained. Drawing 3 is the flow chart which showed the flow of the processing at the time of speech recognition.
[0037] In speech recognition, first, in step S21, the voice for recognition is incorporated in the voice input section 101, and A/D conversion is performed. Step S22 performs sonagraphy in the acoustical-treatment section 102. The analysis method generally used for speech recognition can be used for sonagraphy for a mel cepstrum, a delta mel cepstrum, etc. like the time of initialization. In the time of speech recognition, the processing change-over section 103 connects the acoustical-treatment section 102 and the coding section 106. Therefore, in step S23, the coding section 106 encodes the multi-dimension characteristic quantity parameter for which it asked at step S22 using the clustering resulting table recorded on the voice communication attaching part 105. That is, scalar quantity child-ization for every dimension is performed in the coding section 106.
[0038] The data of each dimension are changed into 4 bits (16step) data by coding. For example, the amount of data in case the data whose analysis period the data of-dimensional [13] and each dimension is 10ms, i.e., per second 100 frames, in 4 bits for the number of dimension of a parameter are transmitted is 13 (dimension). It is set to x4 (bit) x 100 (frame/s) = 5.2kbps.
[0039] Next, sending out and reception of coded data are performed at steps S24 and S25. If in charge of data transfer, the communications control section 107 in the terminal section 100, a communication line 300, and the communications control section 201 of the server section 200 are used as mentioned above. In a communication line 300, the various means of communications which can transmit data are available irrespective of a cable and wireless.
[0040] In the time of speech recognition processing, the processing change-over section 202 connects the communications control section 201 and the decryption section 204. Therefore, the decryption section 204 decodes the multi-dimension characteristic quantity

parameter received in the communications control section 201 using the clustering resulting table recorded on the voice communication information attaching part 203 at step S26. A sound parameter is obtained by this decryption. At step S27, speech recognition is performed using the parameter decoded at step S26. The speech recognition section 205 performs this speech recognition using the sound model held at the sound model attaching part 206. However, unlike general speech recognition, there is no acoustical-treatment section. This is because the data decoded in the decryption section 204 are the sound parameter itself. Moreover, HMM (Hidden Markov Model) is used as a sound model. At step S28, application 207 is operated using the speech recognition result obtained by the speech recognition of step S27. Application 207 is not cared about whether it is in the terminal section 100 even if it is in the server section 200, or both distribute. When application 207 is in the terminal section 100, or when distributing, it is necessary to use the communications control sections 107 and 201 and a communication line 300, and to transmit a recognition result, the data of the internal state of application, etc.

[0041] Since coding and a decryption are performed using a table (clustering resulting table) according to the 1st operation gestalt in order to perform coding which was adapted for the sound condition in coding processing of a sound parameter as explained above, suitable coding is attained to change of the acoustical description. For this reason, decline in the recognition rate by change of an environmental sound can be prevented.

[0042] With the 1st operation gestalt of the <2nd operation gestalt>, the coding conditions (clustering resulting table) adaptation-ized in the sound condition were created, this coding condition was shared between the coding section 106 and the decryption section 204, coding processing decryption processing was performed, and transmission of suitable voice data and speech recognition processing were realized. The 2nd operation gestalt explains how to recognize without decrypting the data further encoded for improvement in the speed of processing.

[0043] Drawing 4 is the block diagram having shown the speech recognition structure of a system concerning the 2nd operation gestalt. Moreover, drawing 5 and drawing 6 are the flow charts explaining actuation of the voice recognition system indicated by the block diagram of drawing 4 . Hereafter, including an example of operation, drawing 4 , drawing 5 , and drawing 6 are related, and it explains. In drawing 4 , the same reference number is given to the same thing as the configuration shown with the 1st operation gestalt. As shown in this drawing, the terminal section 100 is the same as the configuration of the 1st operation gestalt. In the server section 500, the processing change-over section 502 connects the communications control section 201 and the likelihood information generation section 503 at the time of initialization, and connects the communications control section 201 and the speech recognition section 505 at the time of speech recognition processing.

[0044] 503 is the likelihood information generation section and generates likelihood information based on the inputted clustering resulting table and the sound table held at the sound model attaching part 506. Speech recognition can be carried out without decoding coded data using the likelihood information generated here. About likelihood information and its generation, it mentions later. 504 is a likelihood information attaching part and holds the likelihood information generated in the induction information generation section 503. In addition, as a record medium for making likelihood information hold in the likelihood information attaching part 504, various things, such as memory, such as RAM, a floppy disk (FD), and a hard disk (HD), can be used.

[0045] 505 is the speech recognition section and is equipped with the likelihood count section 508 and the language retrieval section 509. Although later mentioned about actuation of the speech recognition section 505, recognition processing is performed to the coded data inputted from the communications control section 201 using the likelihood information held at the likelihood information attaching part 504.

[0046] Hereafter, with reference to drawing 5 and drawing 6 , the speech recognition processing by the 2nd operation gestalt is explained.

[0047] First, initial setting is performed before speech recognition initiation. Initial setting is for performing adaptation by the sound environment of coded data like the 1st operation gestalt. Although speech recognition is possible by using a default about coded data even if this initial setting does not carry out, improvement in a recognition rate is expected by performing initial setting.

[0048] Since it is completely the same as that of the 1st operation gestalt (steps S1-S6) about each processing of steps S40-S45 in the terminal section 100, explanation is omitted. Hereafter, the initialization process of the server section 500 is explained.

[0049] Step S46 receives first the voice communication information (this example clustering resulting table) generated in the terminal section 100 in the communications control section 501. The processing change-over section 502 has connected the likelihood information generation section 503 side in the initialization process. Therefore, likelihood information is generated in step S47. Hereafter, generation of likelihood information is described. Generation of likelihood information is performed in the likelihood information generation section 503 using the sound model currently held at the sound model attaching part 506. This sound model is expressed by HMM etc.

[0050] Although there are various things in the generation method of likelihood information, how to use scalar quantity child-ization here is described. As the 1st operation gestalt explained, the clustering resulting table for the formation of a scalar quantity child for every dimension of a multi-dimension sound parameter is obtained by processing of the terminal section 100 of steps S40-S45. The value and sound model of each quantizing point which are held at this table perform a part of likelihood count about each quantizing point. This value is held to the likelihood information attaching part 504. A decryption becomes unnecessary in order for referring to the table to perform likelihood count based on the scalar quantity child-ized value received as coded data at the time of recognition.

[0051] The approach of the likelihood count by such referring to the table is detailed to the new spring Acoustical Society of Japan lecture collected works 1-5-12 in the Heisei 8 fiscal year "high speed [in speech recognition]" by reference, Sagayama and others. Moreover, the other vector-quantizing method of the formation of a scalar quantity child, the method of carrying out a mixed distribution operation beforehand at each dimension, and omitting addition, etc. can be used. It is introduced to the above-mentioned reference also about these. The above-mentioned count result is held in the form of a table over a scalar quantity child-ized value in step S48 at the likelihood attaching part 504.

[0052] Next, with reference to drawing 6 , it explains that the speech recognition processing by the 2nd operation gestalt flows. Since the processing shown in steps S60-S64 in the terminal section 100 is the same as that of the 1st operation gestalt (steps S20-S24), explanation is omitted.

[0053] Step S65 receives the coded data of the multi-dimension sound parameter required in processing of steps S20-S24 in the communications control section 501 of the server section 500. The processing change-over section 502 is connected the likelihood count section 508 side at the time of speech recognition. The speech recognition section 505 can be divided into the likelihood count section 508 and the language retrieval section 509, and can be expressed. Although likelihood is calculated by the likelihood count section 508, at step S66, it calculates in this case not using a sound model but using the data held at the likelihood information attaching part 504 by performing "refer to the table to a scalar quantity child-ized value." Since the detail of count is indicated in detail by the above-mentioned reference, explanation is omitted here.

[0054] At step S67, to the result of the likelihood count by step S66, language retrieval is performed and a recognition result is obtained. A word dictionary, network syntax, a language model like n-gram, etc. perform language retrieval using the syntax generally used for speech recognition. At step S68, application 507 is operated using the obtained recognition result. In addition, like the 1st operation gestalt, application 507 is not cared about, whether it is in the terminal section 100 even if it is in the server section 500, or both distribute. When application is in the terminal section 100, or when distributing, it is necessary to use the communications control sections 407 and 501 and a communication line 300, and to transmit a recognition result, the data of the internal state of application, etc.

[0055] As mentioned above, since speech recognition can be carried out according to the 2nd operation gestalt, without decrypting the encoded data, improvement in the speed of processing can be attained.

[0056] Speech recognition processing of the 1st and 2nd operation gestalt in which it explained above can be used about the general application using speech recognition. Especially, using a small personal digital assistant as the terminal section 100, when voice input performs control of a device and information retrieval, it is suitable.

[0057] Moreover, according to each above-mentioned operation gestalt, when making a device distribute speech recognition processing and making it operate using coding for speech recognition, coding processing will be performed according to the property of a background noise, internal noise, or a microphone etc. For this reason, when the properties of the bottom of a noise environment or a microphone differ, while being able to prevent degradation of a recognition rate, efficient coding is attained and the merit of being able to press down the transmission amount of data of a channel is obtained.

[0058] In addition, it cannot be overemphasized by the purpose of this invention supplying the storage which recorded the program code of the software which realizes the function of the operation gestalt mentioned above to a system or equipment, and carrying out read-out activation of the program code with which the computer (or CPU and MPU) of the system or equipment was stored in the storage that it is attained.

[0059] In this case, the function of the operation gestalt which the program code itself read from the storage mentioned above will be realized, and the storage which memorized that program code will constitute this invention.

[0060] As a storage for supplying a program code, a floppy disk, a hard disk, an optical disk, a magneto-optic disk, CD-ROM, CD-R, a magnetic tape, the memory card of a non-volatile, ROM, etc. can be used, for example.

[0061] Moreover, it cannot be overemphasized that it is contained also when the function of the operation gestalt which performed a part or all of processing that OS (operating system) which is working on a computer is actual, based on directions of the program code, and the function of the operation gestalt mentioned above by performing the program code which the computer read is not only realized, but was mentioned above by the processing is realized.

[0062] Furthermore, after the program code read from a storage is written in the memory with which the functional expansion unit connected to the functional add-in board inserted in the computer or a computer is equipped, it cannot be overemphasized that it is contained also when the function of the operation gestalt which performed a part or all of processing that CPU with which the functional add-in board and functional expansion unit are equipped based on directions of the program code is actual, and mentioned above by the processing is realized.

[0063]

[Effect of the Invention] As explained above, according to this invention, suitable coding according to change of the acoustical description is attained, and decline in the recognition rate by change of an environmental sound and decline in the compressibility in coding can be prevented.

[Translation done.]

CLAIMS

[Claim(s)]
[Claim 1] An input means to input sound information, and an analysis means to analyze the sound information inputted with said input means, and to acquire a characteristic quantity parameter, A setting means to set up the processing conditions for compression coding based on the characteristic quantity parameter obtained with said analysis means, The conversion means which carries out compression coding of the characteristic quantity parameter obtained with said analysis means according to said processing conditions, The voice recognition system characterized by having the recognition means which carries out speech recognition based on the processing conditions set up with said setting means, and the characteristic quantity parameter by which compression coding was carried out with said conversion means.
[Claim 2] The voice recognition system according to claim 1 characterized by to have further a notice means notify the setups which consisted of the 1st equipment which has said analysis means, said setting means, and said conversion means, and the 2nd equipment which has said recognition means, and were set up with said setting means, and the data acquired with said conversion means from said 1st equipment to said 2nd equipment.
[Claim 3] Said recognition means is a voice recognition system according to claim 1 or 2 characterized by performing speech recognition processing based on the characteristic quantity parameter which was equipped with a decode means to decode said characteristic quantity parameter by which compression coding was carried out with reference to said processing conditions, and was decoded with said decode means.
[Claim 4] The voice recognition system according to claim 2 characterized by equipping said 2nd equipment with a maintenance means to hold said processing conditions notified with said notice means, further.
[Claim 5] Said recognition means is a voice recognition system according to claim 1 or 2 characterized by having said processing conditions, a count means to perform a part of likelihood count in connection with speech recognition based on a sound model, and a means to perform likelihood count to the data acquired with said conversion means using the count result by said count means, and to obtain a speech recognition result.
[Claim 6] The voice recognition system according to claim 5 characterized by having further a maintenance means to hold the count result computed with said count means.
[Claim 7] Said conversion means is a voice recognition system according to claim 1 to 6 characterized by forming into a scalar quantity child the multi-dimension vocal parameter obtained by said analysis means for every dimension.
[Claim 8] The voice recognition system according to claim 7 characterized by using a LBG algorithm in said formation of a scalar quantity child.
[Claim 9] The voice recognition system according to claim 7 which assumes that the data for quantization carry out Gaussian distribution, and a quantization step is this distribution and is characterized by carrying out a quantum so that it may become same probability in said formation of a scalar quantity child.
[Claim 10] Said setting means is a voice recognition system according to claim 7 to 9 characterized by making the clustering for said formation of a scalar quantity child change based on the characteristic quantity parameter obtained with said analysis means.
[Claim 11] The input process which inputs sound information, and the analysis process which analyzes the sound information inputted at said input process, and acquires a characteristic quantity parameter, The setting process which sets up the processing conditions for compression coding based on the characteristic quantity parameter obtained at said analysis process, The conversion process which carries out compression coding of the characteristic quantity parameter obtained at said analysis process according to said processing conditions, The speech recognition approach characterized by having the recognition process which carries out speech recognition based on the processing conditions set up at said setting process, and the characteristic quantity parameter by which compression coding was carried out at said conversion process.
[Claim 12] The speech-recognition approach according to claim 11 characterized by to have further the notice process which notifies the setups which consisted of the 1st equipment which has said analysis process, said setting process, and said conversion process, and the 2nd equipment which has said recognition process, and were set up at said setting process, and the data acquired at said conversion process to said 2nd equipment from said 1st equipment.
[Claim 13] Said recognition process is the speech recognition approach according to claim 11 or 12 characterized by performing speech recognition processing based on the characteristic quantity parameter which was equipped with the decode process which decodes said characteristic quantity parameter by which compression coding was carried out with reference to said processing conditions, and was decoded at said decode process.
[Claim 14] The speech recognition approach according to claim 12 characterized by equipping said 2nd equipment with the maintenance process which holds in memory said processing conditions notified at said notice process further.
[Claim 15] Said recognition process is the speech recognition approach according to claim 11 or 12 characterized by having said processing conditions, the count process which performs a part of likelihood count in connection with speech recognition based on a sound model, and the process which performs likelihood count to the data acquired at said conversion process using the count result by

said count process, and obtains a speech recognition result.

[Claim 16] The speech recognition approach according to claim 15 characterized by having further the maintenance process which holds in memory the count result computed at said count process.

[Claim 17] Said conversion process is the speech recognition approach according to claim 11 to 16 characterized by forming into a scalar quantity child the multi-dimension vocal parameter obtained according to said analysis process for every dimension.

[Claim 18] The speech recognition approach according to claim 17 characterized by using a LBG algorithm in said formation of a scalar quantity child.

[Claim 19] The speech recognition approach according to claim 17 which assumes that the data for quantization carry out Gaussian distribution, and a quantization step is this distribution and is characterized by carrying out a quantum so that it may become same probability in said formation of a scalar quantity child.

[Claim 20] Said setting process is the speech recognition approach according to claim 17 to 19 characterized by making the clustering for said formation of a scalar quantity child change based on the characteristic quantity parameter obtained at said analysis process.

[Claim 21] An input means to input sound information, and an analysis means to analyze the sound information inputted with said input means, and to acquire a characteristic quantity parameter, A setting means to set up the processing conditions for compression coding based on the characteristic quantity parameter obtained with said analysis means, A 1st notice means to notify the processing conditions set up with said setting means to an external device, The information processor characterized by having the conversion means which carries out compression coding of the characteristic quantity parameter obtained with said analysis means according to said processing conditions, and a 2nd notice means to notify the data obtained with said conversion means to said external device.

[Claim 22] The information processor carry out having a 1st receiving means receive the processing conditions which start compression coding from an external device, the maintenance means make the processing conditions which received with said 1st receiving means hold in memory, the 2nd receiving means receive the data by which compression coding was carried out from said external device, and the recognition means perform the speech recognition to the data received with said 2nd receiving means using the processing conditions held at said maintenance means as the description.

[Claim 23] The input process which inputs sound information, and the analysis process which analyzes the sound information inputted at said input process, and acquires a characteristic quantity parameter, The setting process which sets up the processing conditions for compression coding based on the characteristic quantity parameter obtained at said analysis process, The 1st notice process which notifies the processing conditions set up at said setting process to an external device, The information processing approach characterized by having the conversion process which carries out compression coding of the characteristic quantity parameter obtained at said analysis process according to said processing conditions, and the 2nd notice process which notifies the data obtained at said conversion process to said external device.

[Claim 24] The 1st receiving process which receives the processing conditions which start compression coding from an external device, The maintenance process which makes the processing conditions received at said 1st receiving process hold in memory, The information processing approach characterized by having the 2nd receiving process which receives the data by which compression coding was carried out from said external device, and the recognition process which performs speech recognition to the data received at said 2nd receiving process using the processing conditions held at said maintenance process.

[Claim 25] A program for a computer to realize the speech recognition approach according to claim 11 to 20.

[Claim 26] A program for a computer to realize the information processing approach of a publication to either of claims 23 or 24.

[Claim 27] The storage which stores the program of a publication in either of claims 25 or 26.

[Translation done.]